A note on studentized confidence intervals for the change-point

Marie Hušková*, Claudia Kirch†

October 20, 2008

Abstract

We study an AMOC time series model with an abrupt change in the mean and dependent errors that fulfill certain mixing conditions. It is known how to construct resampling confidence intervals using blocking techniques, but so far no studentizing has been considered. A simulation study shows that we obtain better intervals by studentizing.
When studentizing dependent data, we need to use flat-top kernels for the estimation of the asymptotic variance. It turns out that this estimator possibly taking changes into account behaves much better than the corresponding Bartlett estimator. Since the asymptotic distribution of change-point statistics for time-series depends on this value, having a good estimator under the null as well as alternatives is essential for testing problems.

Keywords: block bootstrap, mixing, flat-top kernel, change in mean

AMS Subject Classification 2000: 62G09, 62G15, 60G10

Acknowledgement: The work of the first author was partly supported by the grants GAČR 201/06/0186 and MSM 02162839.

1 Introduction

Recently a number of papers has been published on possible application of bootstrapping or permutation methods in change-point analysis, confer Hušková [5] for a recent survey. Most of these papers are concerned with obtaining critical values for the corresponding

*Charles University of Prague, Department of Statistics, Sokolovská 83,
CZ – 186 75 Praha 8, Czech Republic; huskova@karlin.mff.cuni.cz
†University of Kaiserslautern, Department of Mathematics, Erwin-Schrödinger-Straße,
D–67 653 Kaiserslautern, Germany; ckirch@mathematik.uni-kl.de

change-point tests. Another important issue in change-point analysis, however, is how to obtain confidence intervals for the change-point. In a previous paper [6] we constructed bootstrap confidence intervals for the change-point in a model with dependent data. However, we did not consider studentizing (also called pivoting) techniques and in fact they have not been used in the independent case either for the construction of confidence intervals for the change-point.

In this paper we prove that the studentized confidence intervals are also asymptotically valid, a simulation study shows that their small sample behavior is in fact better than for the original ones without studentizing. In case of dependent data it is not entirely clear how best to do the studentizing, i.e. which estimators for the variance to use. We follow the approach by Götze and Künsch [4] who showed second-order correctness of the block bootstrap when using specific estimators for the variance. Most importantly they use the natural variance estimator for the block bootstrap (details can be found below), which is closely related to the Bartlett estimator. But second-order correctness is only achieved if for the variance estimator of the original sequence different weights than triangular ones (as in the Bartlett estimator) are used. For example one can use the flat-top kernel estimators as suggested by Politis and Romano [10], for which adaptive bandwidth selection procedures exist (cf. Politis [9]).

As a side result we obtain that those estimators taking possible changes into account under the null as well as under alternatives are consistent, moreover they behave much better than the Bartlett estimator. Since the asymptotic of test statistics for change-point problems for time-series usually depends on this asymptotic variance, it is very important to have good estimators in order to construct asymptotic tests with a reasonable small sample bevavior under the null hypothesis as well as alternatives. The Bartlett estimator is not satisfactory in that respect, confer e.g. the simulation study in Kirch [8]. Therefore this result is very important by itself.

We consider the following At-Most-One-Change (AMOC) location model

$$X(i) = \mu + d\,1_{\{i>m\}} + e(i), \quad 1 \leqslant i \leqslant n, \tag{1.1}$$

where $m = m(n) = \lfloor n\vartheta \rfloor$, $0 < \vartheta < 1$, $d = d_n$ may depend on n. The errors $\{e(i), 1 \leqslant i \leqslant n\}$ are stationary and strong-mixing with a rate specified below,

$$\begin{aligned} &\mathrm{E}\, e(i) = 0, \quad 0 < \sigma^2 = \mathrm{E}\, e(i)^2 < \infty, \quad \mathrm{E}\, |e(i)|^\nu < \infty \text{ for some } \nu > 4, \\ &\sum_{h \geqslant 0} |\gamma_h| < \infty, \quad \text{where } \gamma_h = \mathrm{cov}(e(0), e(h)), \end{aligned} \tag{1.2}$$

and

$$\tau^2 := \gamma_0 + 2 \sum_{h \geqslant 1} \gamma_h < \infty. \tag{1.3}$$

Recall the definition of the fourth order cumulant $\kappa(h,r,s)$ given by

$$\kappa(h,r,s) = \mathrm{E}(e(k)e(k+h)e(k+r)e(k+s)) - (\gamma_h\gamma_{r-s} + \gamma_r\gamma_{h-s} + \gamma_s\gamma_{h-r}).$$

We assume now that

$$\sup_h \sum_{r,s} |\kappa(h,r,s)| < \infty. \tag{1.4}$$

In this paper we are interested in getting approximation of the distribution of the following class of change-point estimators

$$\begin{aligned}\widehat{m}(\gamma) &= \arg\max\{|S_\gamma(k)|; k = 1, \dots, n-1\}\\ &= \min\{k;\ 1 \leqslant k < n,\ S_\gamma(k) \geqslant S_\gamma(j),\ j = 1, \dots, n-1\},\end{aligned} \tag{1.5}$$

where

$$S_\gamma(k) = \left(\frac{n}{k(n-k)}\right)^\gamma \sum_{i=1}^{k} (X(i) - \bar{X}_n), \quad 0 \leqslant \gamma \leqslant \frac{1}{2},$$

and $\bar{X}_n = n^{-1} \sum_{i=1}^{n} X(i)$.

In order to prove validity of the developed bootstrap scheme as well as to obtain the asymptotics under the null hypothesis for the change-point estimator we have to use some results like laws of (iterated) logarithm or large numbers for a triangular array. Therefore we need additionally to assumptions (1.2) and (1.4) the following one for certain $\delta, \Delta > 0$ (in some cases > 2):

($\mathcal{A}$) Let $\{e(i) : i \in \mathbb{Z}\}$ be a strictly stationary sequence with $\operatorname{E} e(0) = 0$. Assume there are $\delta, \Delta > 0$ with

$$\operatorname{E} |e(0)|^{2+\delta+\Delta} \leqslant D_1$$

and

$$\sum_{k=0}^{\infty} (k+1)^{\delta/2} \alpha_n(k)^{\Delta/(2+\delta+\Delta)} \leqslant D_2, \tag{1.6}$$

where $\alpha_n(k)$ is the corresponding strong mixing coefficient, i.e.

$$\alpha_n(k) = \sup_{A,B} |P(A \cap B) - P(A)P(B)|\,,$$

where A and B vary over the σ-fields $\mathcal{A}(e(0), e(-1), \dots)$ respectively $\mathcal{A}(e(k), e(k+1), \dots)$.

In fact this assumption is only needed to obtain a Donsker type central limit theorem for the partial sums of the errors (to derive the asymptotics under the null hypothesis) as well as bounds on higher order moments of certain sums of the observed error sequence. This in turn yields laws of large numbers and laws of (iterated) logarithm. The proofs can easily be adapted to allow for errors that do not fulfill condition ($\mathcal{A}$) but the necessary moment conditions. As an example a large class of linear processes fulfills the assumption. More details can be found in Hušková and Kirch [6].

As in the previous paper [6] we will only consider the case $\gamma = 1/2$ in the following. The results for $0 \leqslant \gamma < 1/2$ can be obtained in a similar way as outlined in Antoch et al. [1]. In the simulation study we will also consider other choices of γ.

In the following $\widehat{m} := \widehat{m}(1/2)$, $S(k) := S_{1/2}(k)$.

2 Flat-top kernel estimators taking possible changes into account

In this section we consider the flat-top kernel estimators for τ^2 as in (1.3) of Politis and Romano [10] but taking possible changes into account.

Precisely consider

$$\widehat{\tau}^2 = \widehat{\tau}^2(\Lambda_n) = \widehat{R}(0) + 2\sum_{k=1}^{\Lambda_n} w(k/\Lambda_n)\widehat{R}(k), \tag{2.1}$$

where for $a > b$ we define $\sum_a^b = 0$ and $\widehat{R}(k) = \frac{1}{n}\sum_{t=1}^{n-k}\widehat{e}(t)\widehat{e}(t+k)$ with

$$\begin{aligned}
&\widehat{e}(t) = X(t) - \bar{X}_{\widehat{m}} 1_{\{t\leqslant \widehat{m}\}} - \bar{X}^0_{\widehat{m}} 1_{\{t>\widehat{m}\}},\\
&\widehat{\mu} = \bar{X}_{\widehat{m}}, \qquad \widehat{d}_n = \bar{X}^0_{\widehat{m}} - \bar{X}_{\widehat{m}}.
\end{aligned}$$

$\bar{X}^0_{\widehat{m}} = \frac{1}{n-\widehat{m}}\sum_{i=\widehat{m}+1}^{n} X_i$, and

$$w(t) = \begin{cases} 1, & |t| \leqslant 1/2,\\ 2(1-|t|), & 1/2 < t < 1,\\ 0, & t \geqslant 1.\end{cases}$$

Concerning the bandwidth Λ_n we assume the following

$$\Lambda_n \to \infty, \qquad \frac{\Lambda_n \log n \log\log n}{n} = o(1). \tag{2.2}$$

If we use a triangular weight function $w(t) = (1-|t|)1_{[-1,1]}$ instead, we obtain the classical Bartlett estimator.

In this paper we need the consistency of this estimator in order to be able to obtain the asymptotic correctness of the studentized confidence intervals in the next section as proposed by Götze and Künsch [4].

Apart from this specific application it is very important to have a good estimator for this value (under the null as well as alternatives) since the performance of change-point tests for dependent data depends crucially on it as can e.g. be seen by the simulation study in Kirch [8]. Therefore the following theorem is of independent interest. Note that it also holds for the null hypothesis and all types of alternatives without restriction.

Theorem 2.1. *Assume that* (1.1) - (1.4) *with* $0 < \vartheta < 1$, $d_n = 0$ *corresponds to the null hypothesis. Moreover let assumption* $(\mathcal{A})$ *be fulfilled for some* $\delta, \Delta > 0$. *Then it holds for a bandwidth* Λ_n *fulfilling* (2.2)

$$\widehat{\tau}^2(\Lambda_n) \xrightarrow{P} \tau$$

This remains true for any bounded weight function $\omega(\cdot)$ *which is continuous at* 0 *with* $\omega(0) = 1$.

As for the Bartlett estimator the choice of Λ_n is crucial for the performance of the estimator. But for the flat-top kernel there exists an adapative selection procedure, for which Politis [9] has shown that the bandwidth chosen in that way captures the theoretically optimal rates very well. This is not true anymore if we use this selection procedure with the Bartlett estimator.

We shortly repeat the selection procedure for the sake of completeness:

Adaptive Selection of the bandwidth Λ: *Let $\widehat{\lambda}$ be the smallest positive integer such that $\left|\widehat{R}(\widehat{\lambda}+k)/\widehat{R}(0)\right| < c\sqrt{\log n/n}$, for $k = 1, \dots, K_n$, where $c > 0$ is a fixed constant, and K_n is a positive, nondecreasing integer-valued function of n such that $K_n = o(\log n)$. Then choose $\widehat{\Lambda}_n = 2\widehat{\lambda}$.*

Remark 2.1. The choice of the parameters c and K_n is left to the practitioner. However, Politis [9] suggests to use $c = 2$ and $K_n = 5$ for the usual sample sizes approximately between 100 and 1000. Our simulations show that these values indeed give good results, that are usually better than for other choices of c and K_n. But different choices give still similar results that are better than what we get using the Bartlett estimator (with a variety of choices for Λ_n).

Unlike the Bartlett the flat-top kernel estimator can be negative, especially if τ is small. In this case the Bartlett usually overestimates τ (while for large values of τ both estimators underestimate it). The flat-top kernel manages to capture small τ quite well and large τ still better than the Bartlett. Nevertheless it is not desirable for many applications that this estimator is too close to zero (in comparison to τ). The reason is that for testing we typically divide by it so that the statistic becomes quite large thus rejecting the null hypothesis wrongly, for confidence intervals it decreases the length of the interval too much.

This is why we propose to use the adapted estimator

$$\widetilde{\tau}^2 = \widetilde{\tau}^2(\Lambda_n) = \max(\widehat{\tau}^2(\Lambda_n), 1/\log^2 n). \tag{2.3}$$

3 Studentized confidence intervals for the change-point estimator

Antoch et al. [1] have used bootstrapping methods to obtain confidence intervals for the change-point in the independent situation, Hušková and Kirch [6] have used a circular overlapping block bootstrap for the dependent situation. Both papers do not consider studentizing although the results for small samples can be significantely improved in this way.

We will concentrate on the dependent situation here, since the extension to the i.i.d. situation is straightforward.

First, we shortly explain how the circular overlapping block bootstrap works, then we introduce studentizing and give the main result showing that the studentized confidence intervals are in fact asymptotically correct.

Let $\widehat{\mu}_1$ be an estimator for μ, $\widehat{\mu}_2$ for $\mu + d_n$, and $\widehat{d_n}$ for d_n satisfying $\widehat{d_n} - d_n = O(\sqrt{\log(n)/n})$ *a.s.*, e.g.

$$\widehat{\mu}_1 = \frac{1}{\widehat{m}} \sum_{i=1}^{\widehat{m}} X(i), \qquad \widehat{\mu}_2 = \frac{1}{n - \widehat{m}} \sum_{i=\widehat{m}+1}^{n} X(i), \qquad \widehat{d_n} = \widehat{\mu}_2 - \widehat{\mu}_1, \tag{3.1}$$

where $\widehat{m} = \widehat{m}(1/2)$ as in (1.5).

Define the estimated residuals and the centered residuals by

$$\widehat{e}(i) = X(i) - \widehat{\mu}_1 1_{\{i \leqslant \widehat{m}\}} - \widehat{\mu}_2 1_{\{i > \widehat{m}\}},$$

$$\widetilde{e}(i) = \widehat{e}(i) - \frac{1}{n} \sum_{j=1}^{n} \widehat{e}(j),$$

respectively.

We will use a circular overlapping block bootstrap of the centered estimated residuals $\{\widetilde{e}(i) : 1 \leqslant i \leqslant n\}$. It has the advantage of being automatically centered around the sample mean.

The general idea of block bootstrapping methods is to split the observation sequence of length n into sequences of length K. Then we put L of them together to a bootstrap sequence (i.e. $n = KL$). We keep the order within the blocks. K and L depend on n and converge to infinity with n.

The idea is that, for a properly chosen block-length K, the block contains enough information about the dependency structure so that the estimate is close to the null hypothesis.

We assume in the following that

$$K, L \to \infty \quad \text{ and } \quad K = K(L),\ n = n(L) = KL,\ K/L = o(1). \tag{3.2}$$

Let $\{U(i) : 1 \leqslant i \leqslant L\}$ be i.i.d. with $P(U(1) = i) = \frac{1}{n}$ for $i = 0, \dots, n-1$ independent of the observations $X(1), \dots, X(n)$. Take the i.i.d. bootstrap sample $e^*(Kl + k) = \widetilde{e}(U(l) + k)$, where $\widetilde{e}(j) = \widetilde{e}(j - n)$ for $j > n$ (hence the name circular bootstrap).

Consider the bootstrap observations

$$X^*(j) = e^*(j) + \widehat{\mu}_1 1_{\{j \leqslant \widehat{m}\}} + \widehat{\mu}_2 1_{\{\widehat{m} < j \leqslant n\}}.$$

We deal now with the following bootstrap estimator of the change-point m

$$\widehat{m}^* = \arg\max\{|S^*(k)|; k = 1, \dots, n-1\}, \tag{3.3}$$

where

$$S^*(k) = \left(\frac{n}{k(n-k)} \right)^{1/2} \sum_{i=1}^{k} (X^*(i) - \bar{X}_n^*).$$

Moreover we also define

$$\widehat{d}_n^* = \frac{1}{n - \widehat{m}^*} \sum_{i=\widehat{m}^*+1}^{n} X^*(i) - \frac{1}{\widehat{m}^*} \sum_{i=1}^{\widehat{m}^*} X^*(i)$$

and

$$\widehat{\tau}^{2*} = \widehat{\tau}^{2*}(K) = \frac{1}{L}\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K}(e^*(Kl+k) - \bar{e}_n^*)\right)^2, \tag{3.4}$$

which is the obvious estimate of the conditional variance of the bootstrap sample.

In our previous paper [6] we showed that it is asymptotically correct to approximate $\widehat{m} - m$ by $\widehat{m}^* - \widehat{m}$, so that the confidence interval for the change-point m can be approximated by the distribution of $2\widehat{m} - \widehat{m}^*$.

In this paper we want to use studentizing meaning that we will now approximate $(\widehat{d}/\widehat{\tau})^2(\widehat{m} - m)$ by $(\widehat{d}^*/\widehat{\tau}^*)^2(\widehat{m}^* - \widehat{m})$. This usually leads to better small sample results since it captures not only first order properties but also second order properties for a large class of statistics. For independent data this is quite straightforward, since we can use $\widehat{\tau}^2 = \frac{1}{n-1}\sum_{i=1}^n(\widetilde{e}(i))^2$ and $\widehat{\tau}^{2*} = \frac{1}{n-1}\sum_{i=1}^n(e^*(i) - \bar{e}_n^*)^2$. However in the dependent situation it is not clear what variance estimators should be used for the original sample as well as the bootstrap sample. In fact, Götze and Künsch [4] investigated for which variance estimators second-order correctness is obtained. Surprisingly it turns out that one needs to use $\widehat{\tau}^{2*}$ as in (3.4), which is closely related to the Bartlett estimator, but that one cannot use the Bartlett for the estimation of τ^2, but has to take an estimator like the flat-top kernel estimator from the previous section, i.e. $\widetilde{\tau}^2$ as in (2.3).

With P^*, E^*, $\mathrm{var}^*, \ldots$ we will denote probability, expectation, variance,..., given $X(1), \ldots, X(n)$.

Theorem 3.1. *Assume that* (1.1) - (1.4) *with* $0 < \vartheta < 1$, (2.2) *and* (3.2) *hold. Moreover let*

$$\widehat{d}_n - d_n = O\left(\sqrt{\frac{\log n}{n}}\right) \qquad P - a.s. \tag{3.5}$$

be fulfilled in addition to

$$d_n^{-2}n^{-1}\log n \to 0 \tag{3.6}$$

as well as $d_n \to 0$ *(i.e. a local change). Moreover let assumption* $(\mathcal{A})$ *be fulfilled for some*

$$0 < \delta^{(2)} + \Delta^{(2)} < (\delta^{(1)} - 2)/2 < (\nu - 4)/2, \quad \Delta^{(1)} = \nu - 2 - \delta^{(1)}. \tag{3.7}$$

If $d_n^2 K \to 0$*, then*

$$\sup_{x\in\mathbb{R}}\left|P^*\left(\frac{\widehat{d}^{*2}}{\widehat{\tau}^{2*}(K)}(\widehat{m}^* - \widehat{m}) \leqslant x\right) - P\left(\frac{\widehat{d}^2}{\widetilde{\tau}^2(K)}(\widehat{m} - m) \leqslant x\right)\right| \to 0 \qquad P - a.s.$$

Since the limit distribution (for both the bootstrap as well as the original statistic under the null) is continuous (as has been pointed out by Remark 2.1 in Hušková and Kirch [6]) the described sampling scheme provides bootstrap approximations to the $(1-\alpha)$-quantile for arbitrary $\alpha \in (0, 1)$. Thus the bootstrap based approximation for the change-point

m can be constructed along the usual lines. Precisely the $(1-\alpha)$-bootstrap confidence interval is given by

$$\left[\widehat{m} - \frac{\widehat{d}_n^{*2}\widetilde{\tau}^2(K)}{\widehat{d}_n^2\widehat{\tau}^{2*}(K)}(q_U^*(\alpha/2)-\widehat{m}), \widehat{m} - \frac{\widehat{d}_n^{*2}\widetilde{\tau}^2(K)}{\widehat{d}_n^2\widehat{\tau}^{2*}(K)}(q_L^*(\alpha/2)-\widehat{m})\right],$$

where

$$q_L^*(\alpha/2) = \sup\{u;\ P^*(\widehat{m}^* < u) \leqslant \alpha/2\}$$

and

$$q_U^*(\alpha/2) = \inf\{u;\ P^*(\widehat{m}^* > u) \leqslant \alpha/2\}.$$

Usually one uses the empirical bootstrap distribution of $\widehat{m}^*$ for say 10 000 random bootstrap samples.

4 Simulations

The simulation study is divided into two parts, in a first part we will investigate how well the new variance estimator based on flat-top kernels behaves, in a second part we investigate the small sample behavior of the studentized confidence intervals. Due to limitations of space and similarity of results we present here only a very small part of the simulations study – the complete simulation results (.pdf-file, 143 pages, 16 MB for the simulation study concerning the flat-top kernel estimators, and 191 pages, 23 MB for the simulation study concerning the studentized confidence intervals) can be obtained from the second author upon request. In the following simulation study we use a slightly different way of estimating the covariances than before, namely we do not use 'mixed terms' (where we use the product of two estimated errors before and after the change). Precisely we use (with appropriate weights – either flat-top or Bartlett)

$$\widehat{\tau}^2 = \widehat{\tau}^2(\Lambda_n) = \widehat{R}(0) + 2\sum_{k=1}^{\Lambda_n} w(k/\Lambda_n)\widehat{R}(k),$$

where

$$\widehat{R}(k) = \frac{1}{n}\left(\sum_{t=1}^{\widehat{m}-k}(X(t)-\bar{X}_{\widehat{m}})(X(t+k)-\bar{X}_{\widehat{m}}) + \sum_{t=\widehat{m}+1}^{n-k}(X(t)-\bar{X}_{\widehat{m}}^0)(X(t+k)-\bar{X}_{\widehat{m}}^0)\right).$$

In simulations (cf. the above mentioned complete simulation results) these estimators behave almost the same as the ones based on the entire estimated error sequence, but sometimes slightly better. However, some additional theoretic problems arise.

4.1 Flat-top kernel estimators

In this subsection we want to investigate how well the flat-top kernel estimator with the automatic bandwidth selection procedure as proposed by Politis [9] works in comparison with the Bartlett estimator. Moreover we will investigate how sensitive the procedure is with respect to the selection of the parameters c and K_n.

To this end we use AR(1) as well as MA(1) sequences with normal as well as exponential errors, change-points at $m = 1/4n, 1/2n, 3/4n$, $n = 80, 200$, and different jump sizes $d = 0, 0.5, 1, 2$. We then plot the estimated density of the estimator (using the standard R routine, which uses a Gaussian kernel, where the bandwidth is chosen according to Silverman's rule of thumb (Silverman [12], p.48, Equation (3.31)). The vertical line indicates the true value which are estimated.

Since the results are very similar no matter what type of errors we use, as well as for different locations and sizes of the change-point, we give only some results for exponential errors, a change-point at $m = 1/4n$ as well as $d = 0, 1$. The complete simulation results can be obtained from the second author (.pdf-file, 143 pages, 20 MB).

The results can be found in Figures 4.1(i) and 4.1(ii).

The estimated density for the Bartlett kernel can be negative (for small values of τ), but this is just an artifact from the density estimation procedure, since the Bartlett estimator is by definition positive. The flat-top estimator on the other hand can become negative (for small values of τ), but this is also the reason why it estimates small values of τ better (confer e.g. Figure 4.1(ii)).

The simulations show that the automatic bandwidth selection procedure is not very sensitive to what initial parameters for c and K_n are used – in fact the results are always very similar. As a contrast the Bartlett estimator depends crucially on a good choice of bandwidth (in the simulations we used the bandwidthes $L = 0.5n, 0.1n, 0.2n$). Moreover the optimal bandwidth depends on what type of dependency is present.

It is clear that the flat-top kernel estimator with automatic bandwidth selection is always at least as good as the Bartlett estimator with the according to the simulations best bandwidth, most of the time even better. Moreover both estimators are significantely underestimating the true value for large τ (confer e.g. Figure 4.1(i), (5) – (8)). However, for larger sample sizes the estimation does get better. The Bartlett estimator additionally overestimates small values of τ, while the flat-top estimaterdoes fine there.

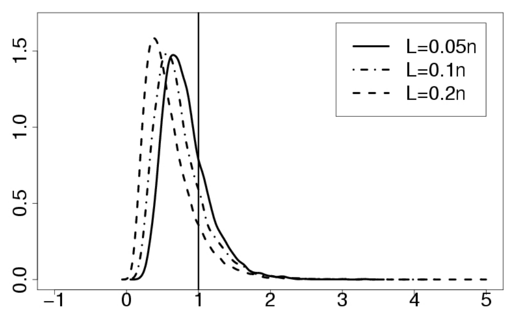
4.2 Studentized confidence intervals

In the previous chapter we have established the asymptotic validity of the studentized bootstrap confidence intervals. The question remains how well these confidence intervals behave for small samples and also how well they behave in comparison with the asymptotic resp. non-studentized intervals.

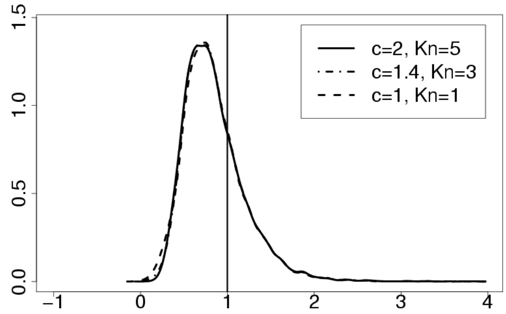
In this section we not only consider $\gamma = \frac{1}{2}$ but also $\gamma = 0$. The difference is that the asymptotic confidence intervals depend on the unknown parameter ϑ for $0 \leqslant \gamma < \frac{1}{2}$.

Moreover we consider changes in the mean of size $d = d_n = 0.5, 1, 2$. The latter ones can hardly be regarded as local changes, however we are still interested in the behavior of the bootstrap intervals, since we conjecture it will also be valid in those cases.

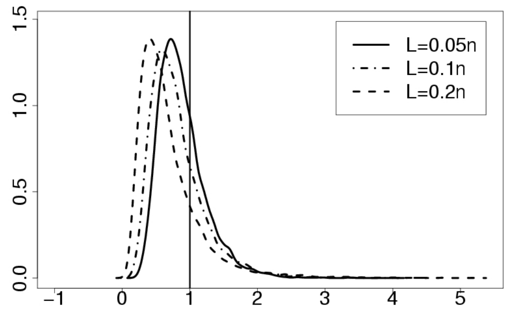
For the simulations we use an autoregressive as well as moving average sequence of order one as an error sequence with standard normally distributed innovations as well as exponential innovations and different values of $\rho = -0.8, -0.3, -0.1, 0.1, 0.3, 0.8$. We consider changes at $m = n/4, n/2$, $n = 80, 200$. For the asymptotic confidence intervals



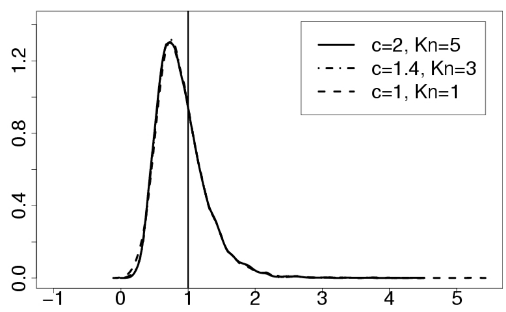
(1) Bartlett: independent errors, $d = 0$ (null), $n = 80$



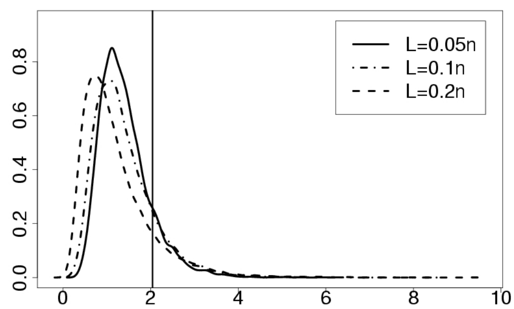
(2) Flat-top: independent errors, $d = 0$ (null), $n = 80$



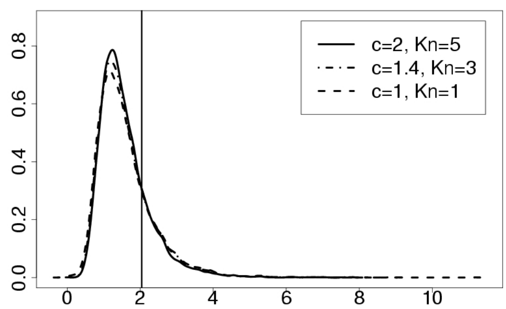
(3) Bartlett: independent errors, $d = 1$, $n = 80$



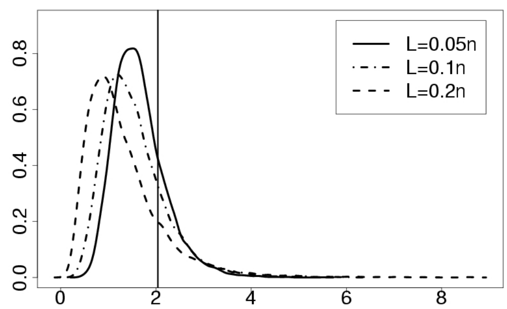
(4) Flat-top: independent errors, $d = 1$, $n = 80$



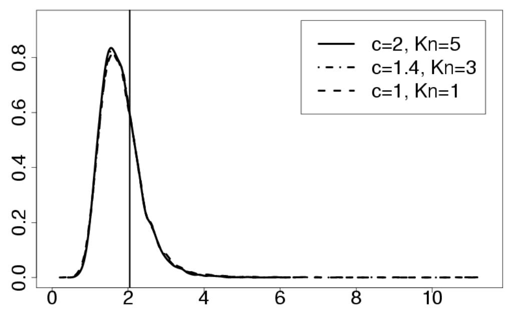
(5) Bartlett: AR(1), $\rho = 0.3$, $d = 1$, $n = 80$



(6) Flat-top: AR(1), $\rho = 0.3$, $d = 1$, $n = 80$

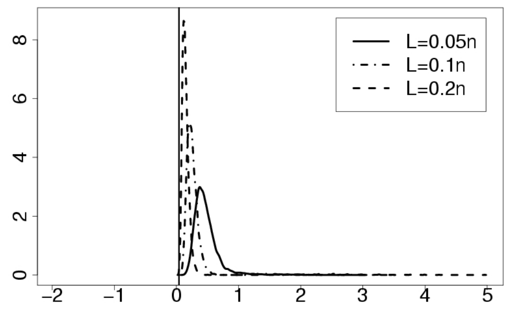


(7) Bartlett: AR(1), $\rho = 0.3$, $d = 1$, $n = 200$

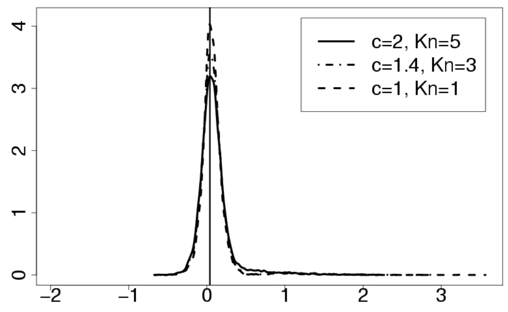


(8) Flat-top: AR(1), $\rho = 0.3$, $d = 1$, $n = 200$

Figure 4.1(i): Comparison of Bartlett estimator with flat-top kernel estimator, change at $m = 1/4n$, exponential errors



(9) Bartlett: MA(1), $\rho = -0.8$, $d = 1$, $n = 80$



(10) Flat-top: MA(1), $\rho = -0.8$, $d = 1$, $n = 80$

Figure 4.1(ii): Comparison of Bartlett estimator with flat-top kernel estimator, change at $m = 1/4n$, exponential errors

we use the flat-top kernel estimator (2.1) with the described automatic bandwidth selection procedure ($K_n = 3$, $c = 1.4$). This yields already better results than using the Bartlett estimator with bandwidth $0.1n$ as done in Hušková and Kirch [6].

For the sake of completeness we also simulate the studentized confidence intervals for the i.i.d. situation. We will not show the results here due to limitations of space, but the studentizing also gives better results in the i.i.d. situation.

A very small selection of representative results is given in this section, the complete simulation results can be obtained from the second author upon request (.pdf-File, 191 pages, 23 MB).

The goodness of confidence intervals can essentially be determined by two criteria:

$\mathcal{C}.1$ The probability that the actual change-point is outside the (1-α)-confidence interval should be close to (smaller than) α.

$\mathcal{C}.2$ The confidence intervals should be short.

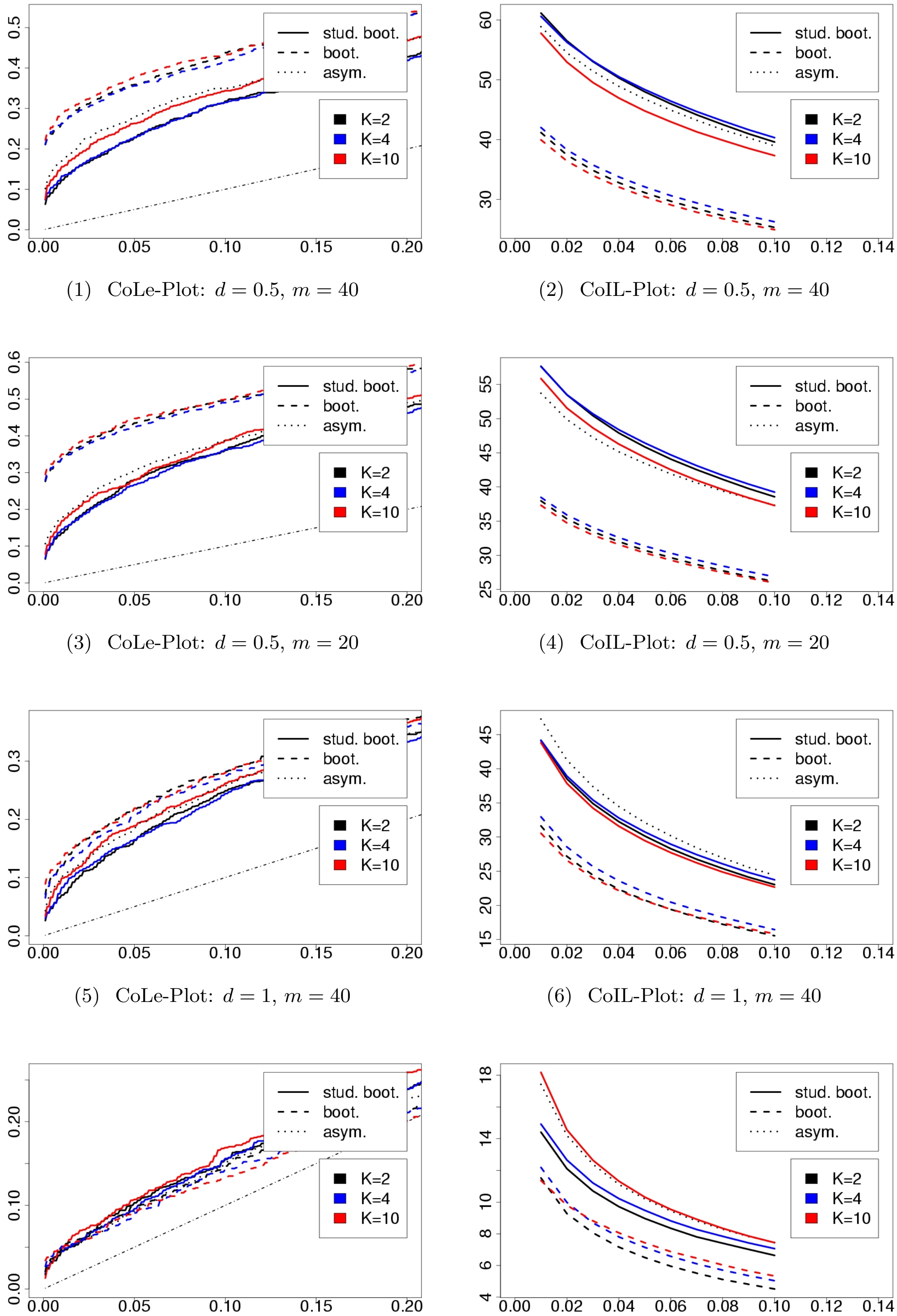
We visualize the first quantity by using CoLe-Plots (**Co**nfidence-**Le**vel-Plots) and the second one by using CoIL-Plots (**Co**nfidence-**I**nterval-**L**ength-Plots).

CoLe-Plots

We explain how the plots are created using the example of asymptotic confidence intervals. The general version of Theorem 2.1 in Hušková and Kirch [6] yields that the asymptotic confidence intervals are calculated using the distribution of $Z = \widehat{m} - \widehat{\tau}^2/\widehat{d_n^2} V(\widehat{\vartheta}, \gamma)$, where $V(\widehat{\vartheta}, \gamma)$ is as in Remark 2.4 in Hušková and Kirch [6] and $\widehat{\vartheta} = \widehat{m}/n$. The CoLe-Plots now draw the empirical distributions function (based on 1 000 observation sequences) of $2\min(P(Z \leqslant m), P(Z \geqslant m))$. This visualizes $\mathcal{C}.1$, since

$$\begin{aligned} m \notin CI(1-\alpha) \quad &\Longleftrightarrow \quad P(Z \leqslant m) \leqslant \alpha/2 \quad \text{or} \quad P(Z \geqslant m) \leqslant \alpha/2 \\ &\Longleftrightarrow \quad 2\min(P(Z \leqslant m), P(Z \geqslant m)) \leqslant \alpha. \end{aligned}$$

Thus for given α on the x-axis the plot shows the empirical probability that m is outside



(1) CoLe-Plot: $d = 0.5$, $m = 40$

(2) CoIL-Plot: $d = 0.5$, $m = 40$

(3) CoLe-Plot: $d = 0.5$, $m = 20$

(4) CoIL-Plot: $d = 0.5$, $m = 20$

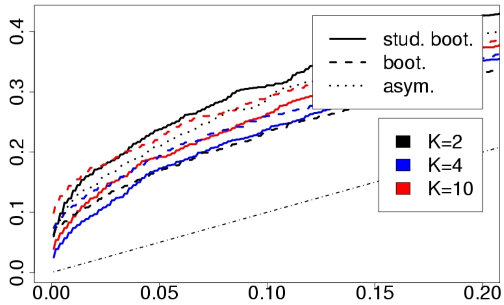
(5) CoLe-Plot: $d = 1$, $m = 40$

(6) CoIL-Plot: $d = 1$, $m = 40$

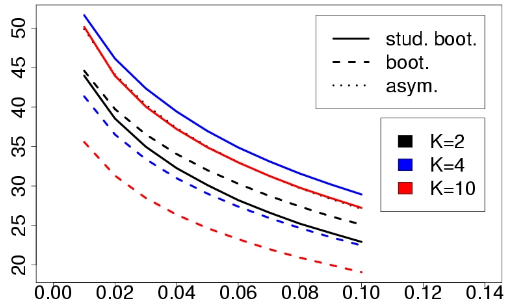
(7) CoLe-Plot: $d = 2$, $m = 40$

(8) CoIL-Plot: $d = 2$, $m = 40$

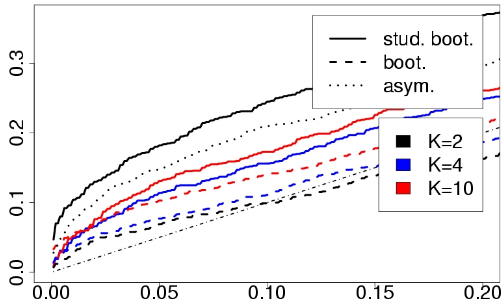
Figure 4.2: Comparison of different confidence intervals, AR(1) process with normal errors, $\rho = 0.3$, $\gamma = 0$, $n = 80$



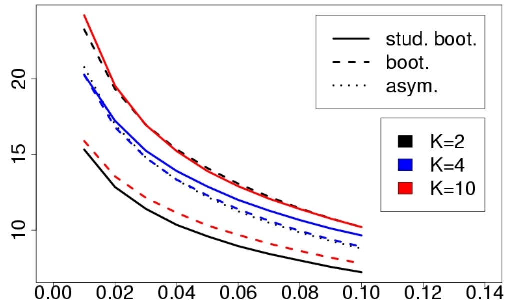
(1) CoLe-Plot: $\rho = -0.3$, $d = 0.5$



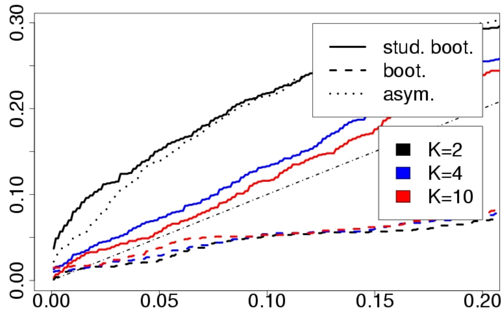
(2) CoIL-Plot: $\rho = -0.3$, $d = 0.5$



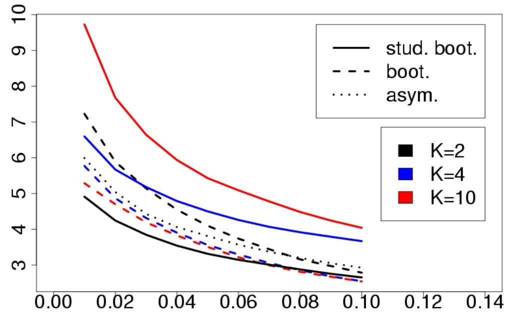
(3) CoLe-Plot: $\rho = -0.3$, $d = 1$



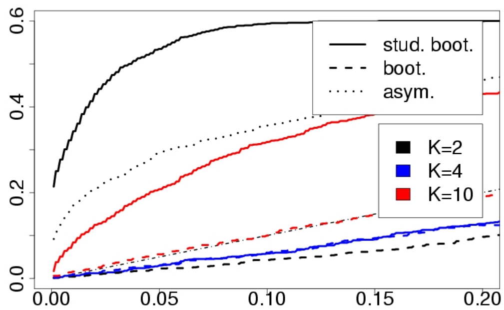
(4) CoIL-Plot: $\rho = -0.3$, $d = 1$



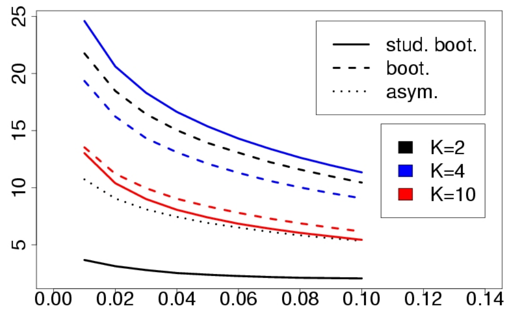
(5) CoLe-Plot: $\rho = -0.3$, $d = 2$



(6) CoIL-Plot: $\rho = -0.3$, $d = 2$

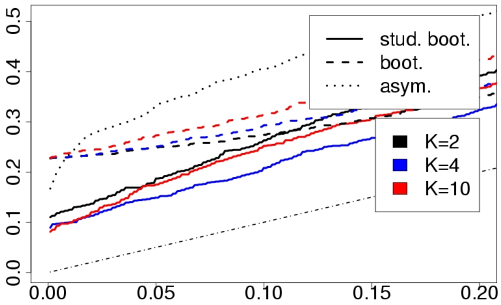


(7) CoLe-Plot: $\rho = -0.8$, $d = 1$

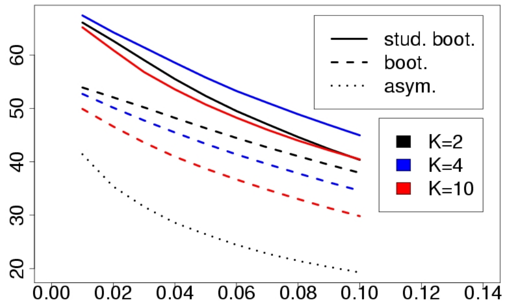


(8) CoIL-Plot: $\rho = -0.8$, $d = 1$

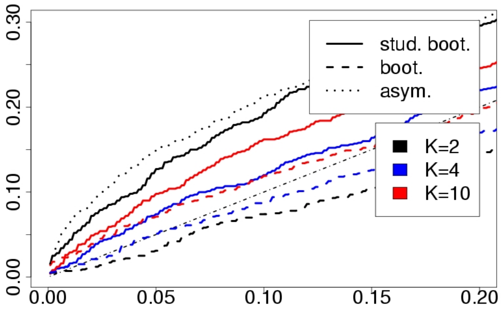
Figure 4.3: Comparison of different confidence intervals, AR(1) process with normal errors, $\gamma = 0$, $n = 80$, $m = 40$



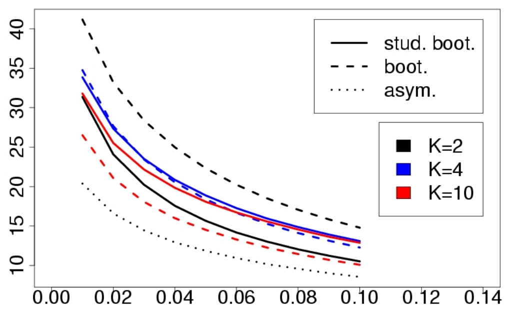
(1) CoLe-Plot: $\rho = -0.3$, $d = 0.5$



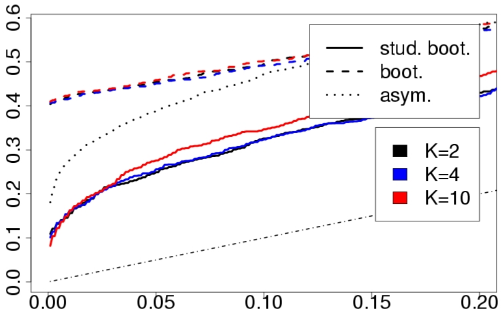
(2) CoIL-Plot: $\rho = -0.3$, $d = 0.5$



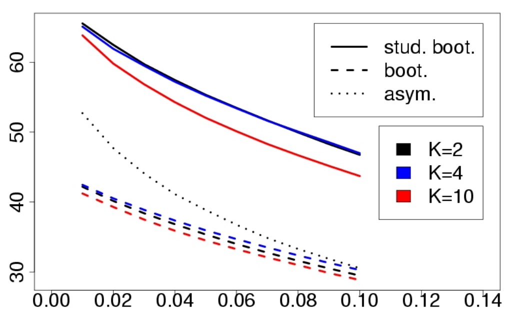
(3) CoLe-Plot: $\rho = -0.3$, $d = 1$



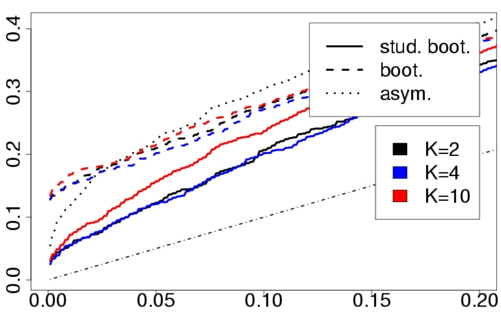
(4) CoIL-Plot: $\rho = -0.3$, $d = 1$



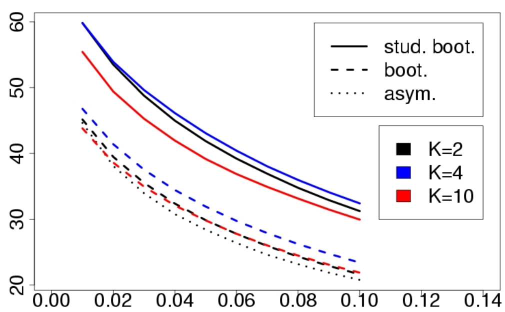
(5) CoLe-Plot: $\rho = 0.3$, $d = 0.5$



(6) CoIL-Plot: $\rho = 0.3$, $d = 0.5$

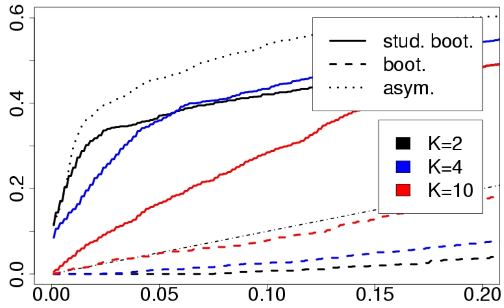


(7) CoLe-Plot: $\rho = 0.3$, $d = 1$

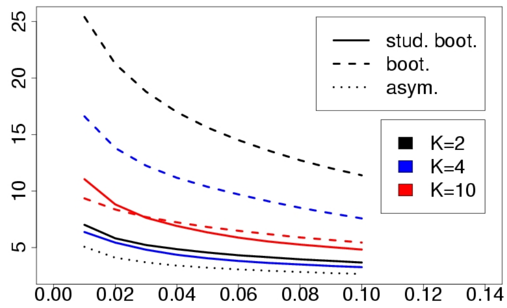


(8) CoIL-Plot: $\rho = 0.3$, $d = 1$

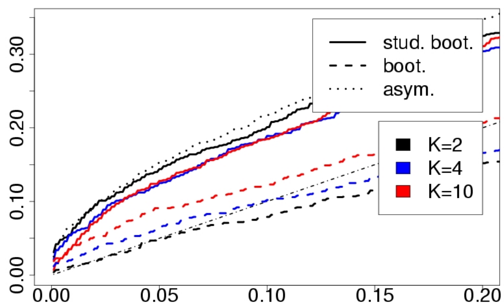
Figure 4.4: Comparison of different confidence intervals, AR(1) process with normal errors, $n = 80$, $m = 40$, $\gamma = 0.5$



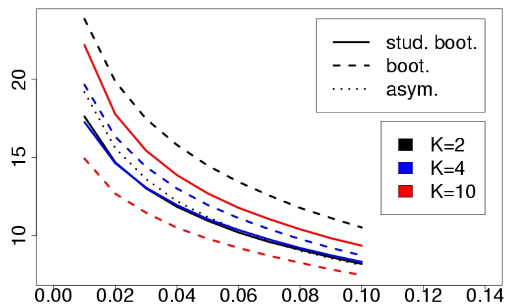
(1) CoLe-Plot: $\rho = -0.8$



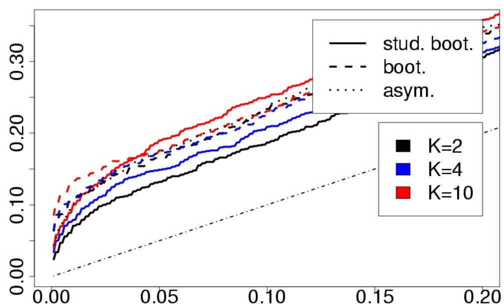
(2) CoIL-Plot: $\rho = -0.8$



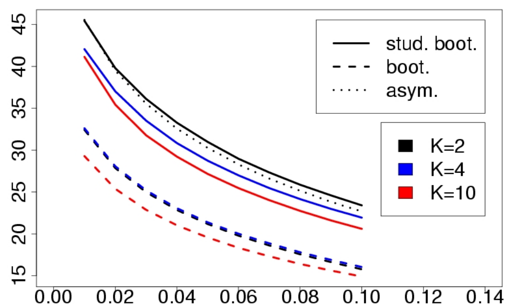
(3) CoLe-Plot: $\rho = -0.3$



(4) CoIL-Plot: $\rho = -0.3$



(5) CoLe-Plot: $\rho = 0.3$



(6) CoIL-Plot: $\rho = 0.3$

Figure 4.5: Comparison of different confidence intervals, MA(1) process with normal errors, $n = 80$, $m = 40$, $d = 1$

the $(1-\alpha)$-confidence interval on the y-axis, hence it visualizes $\mathcal{C}.1$. Optimally, the plot should be below or (even better) on the diagonal.

For the bootstrap confidence intervals the procedures works exactly the same but now the intervals are calculated using the (empirical, based on 10 000 resamples) distribution of $\widetilde{Z} = \widehat{m} - (\widehat{d}_n^* \widehat{\tau}/(\widehat{\tau}^* \widehat{d}_n)^2 (\widehat{m}^* - \widehat{m})$.

CoIL-Plots

We calculate the length of the confidence intervals for levels $\alpha = 0.01, 0.02, \ldots, 0.1$ based on 1 000 observation sequences. The empirical bootstrap distribution is based on 10 000 random samples as before. Then we plot the mean using a line, linearly interpolated. So these plots visualize the length of the intervals and thus $\mathcal{C}.2$. However, since the methods above can yield confidence intervals $[l_\alpha, u_\alpha]$ with $l_\alpha < 1$ or $u_\alpha > n$, we correct the length correspondingly, i.e. consider the intervals $[\max(1, l_\alpha), \min(n, u_\alpha)]$. This is different from the CoIL-Plots in Hušková and Kirch [6].

Note that the scale on the y-axis is not the same for different pictures. This way we can better compare different methods.

The results of the simulation can be found in Figures 4.2, 4.3, 4.4 and 4.5.

Most of the time the studentized bootstrap behaves best in terms of the coverage probability $\mathcal{C}.1$, in principal it works better for a positive correlation than for a negative one. It is interesting that under strong negative correlations the choice of block-length is crucial for the performance, see especially 4.4 (7), while this is not the case for the non-studentized bootstrap. This is probably due to the fact that the performance of the variance estimator is much more important for the studentized than the non-studentized bootstrap. Moreover, for a MA(1) sequence with $\rho = -0.8$ the studentized bootstrap generally does not work anymore. The reason probably is (confer also Figure 4.1(ii)) that the confidence interval is too small. In the bootstrap we multiply with the variance from the flat-top kernel, which is rather small, and divide by the bootstrap variance (close to the Bartlett variance estimator), which is too large.

5 Proofs

Now we will prove our first theorem.

Proof of Theorem 2.1. First we consider the subsequence with $\widehat{m} < m$. We prove that we can replace the estimated error sequence in the definition of $\widehat{R}(k)$ by the actual errors $\{e(i)\}$. First note that

$$\widehat{e}(i) = \begin{cases} e(i) - \bar{e}_{\widehat{m}}, & i \leqslant \widehat{m}, \\ e(i) - \bar{e}^0_{\widehat{m}} + d_n \left(1_{\{i>m\}} - \frac{n-m}{n-\widehat{m}}\right), & i > \widehat{m}. \end{cases}$$

Recall

$$\begin{aligned} \widehat{R}(k) &= \frac{1}{n}\sum_{t=1}^{n-k} \widehat{e}(t)\widehat{e}(t+k) \\ &= \frac{1}{n}\sum_{t=1}^{n-k}(e(t) - \bar{e}_{\widehat{m}}1_{\{t\leqslant\widehat{m}\}} - \bar{e}^0_{\widehat{m}}1_{\{t>\widehat{m}\}})(e(t+k) - \bar{e}_{\widehat{m}}1_{\{t+k\leqslant\widehat{m}\}} - \bar{e}^0_{\widehat{m}}1_{\{t+k>\widehat{m}\}}) \\ &\quad + d_n\frac{1}{n}\sum_{t=\max(\widehat{m}-k,0)+1}^{\min(\widehat{m},n-k)}(e(t) - \bar{e}_{\widehat{m}})\left(1_{\{t>m-k\}} - \frac{n-m}{n-\widehat{m}}\right) \\ &\quad + d_n\frac{1}{n}\sum_{t=\widehat{m}+1}^{n-k}(e(t) - \bar{e}^0_{\widehat{m}})\left(1_{\{t>m-k\}} - \frac{n-m}{n-\widehat{m}}\right) \\ &\quad + d_n\frac{1}{n}\sum_{t=\widehat{m}+1}^{n-k}(e(t+k) - \bar{e}^0_{\widehat{m}})\left(1_{\{t>m\}} - \frac{n-m}{n-\widehat{m}}\right) \end{aligned}$$

$$+ d_n^2 \frac{1}{n} \sum_{t=\widehat{m}+1}^{n-k} \left(1_{\{t>m\}} - \frac{n-m}{n-\widehat{m}}\right)\left(1_{\{t>m-k\}} - \frac{n-m}{n-\widehat{m}}\right)$$
$$= A_1(k) + A_2(k) + A_3(k) + A_4(k) + A_5(k).$$

By an application of the Hájek -Renyi inequality (cf. Kirch [7], Theorem B.3)

$$\max_{1\leqslant k\leqslant n} \left| \frac{1}{\sqrt{k}} \sum_{i=1}^{k} e(i) \right| = O_P(\sqrt{\log n}) \tag{5.1}$$

From this we can conclude

$$A_1(k) = \frac{1}{n}\sum_{t=1}^{n-k} e(t)e(t+k) + O_P\left(\frac{\log n}{n}\right),$$

where the last term is uniformly in k. From Berkes et al. [3], Theorem 1.1 (i) (precisely equation (2.1) in the proof) and Remark 1.1 we obtain

$$A_1(0) + 2\sum_{k=1}^{\Lambda_n} w(k/\Lambda_n)A_1(k) = \tau^2 + o_P(1).$$

By Theorem 2.1 a) and Remark 2.1 in Hušková and Kirch [6] we obtain

$$\widehat{m} - m = O_P(d_n^{-2}) \tag{5.2}$$

if $d_n^{-2}n^{-1}\log n \to 0$.

By (5.1) and (5.2) uniformly in k

$$\begin{aligned}
&A_2(k)\\
&= d_n \frac{1}{n} \sum_{t=\max(\min(m-k,\widehat{m}),0)+1}^{\min(\widehat{m},n-k)} \frac{m-\widehat{m}}{n-\widehat{m}}(e(t)-\bar{e}_{\widehat{m}}) - d_n\frac{1}{n}\sum_{t=\max(\widehat{m}-k,0)+1}^{\min(m-k,\widehat{m})} \frac{n-m}{n-\widehat{m}}(e(t)-\bar{e}_{\widehat{m}})\\
&= O_P\left(|d_n|\frac{\sqrt{(m-\widehat{m})\log n}}{n}\right) = \begin{cases} O_P\left(\frac{\sqrt{\log n}}{n}\right), & d_n \geqslant \sqrt{\frac{\log n \log\log n}{n}},\\ O_P(d_n(\log n/n)^{1/2}), & \text{else}\end{cases}\\
&= O_P\left(\frac{\log n\log\log n}{n}\right).
\end{aligned}$$

Similarly

$$\begin{aligned}
A_3(k) &= d_n\frac{1}{n}\sum_{t=\max(m-k,\widehat{m})+1}^{n-k} \frac{m-\widehat{m}}{n-\widehat{m}}(e(t)-\bar{e}^0_{\widehat{m}}) - d_n\frac{1}{n}\sum_{t=\widehat{m}+1}^{m-k}\frac{n-m}{n-\widehat{m}}(e(t)-\bar{e}^0_{\widehat{m}})\\
&= O_P\left(|d_n|\frac{\sqrt{(m-\widehat{m})\log n}}{n}\right) = \begin{cases} O_P\left(\frac{\sqrt{\log n}}{n}\right), & d_n \geqslant \sqrt{\frac{\log n \log\log n}{n}},\\ O_P(d_n(\log n/n)^{1/2}), & \text{else}\end{cases}\\
&= O_P\left(\frac{\log n\log\log n}{n}\right),\\
A_4(k) &= O_P\left(\frac{\log n\log\log n}{n}\right),
\end{aligned}$$

and

$$\begin{aligned}
&A_5(k) \\
&= \frac{d_n^2}{n} \sum_{t=\widehat{m}+1}^{m-k} \left(\frac{n-m}{n-\widehat{m}} \right)^2 - \frac{d_n^2}{n} \sum_{t=\max(m-k,\widehat{m})+1}^{m} \frac{(n-m)(m-\widehat{m})}{(n-\widehat{m})^2} \\
&\quad + \frac{d_n^2}{n} \sum_{t=m+1}^{n} \left(\frac{m-\widehat{m}}{n-\widehat{m}} \right)^2 \\
&= O\left(\frac{d_n^2}{n}(m-\widehat{m}) \right) = \begin{cases} O_P\left(\frac{1}{n}\right), & d_n \geqslant \sqrt{\frac{\log n \log\log n}{n}}, \\ O(d_n^2), & \text{else} \end{cases} \\
&= O_P\left(\frac{\log n \log\log n}{n} \right).
\end{aligned}$$

Putting everything together we get

$$\begin{aligned}
&\widehat{\tau}^2(\Lambda_n) = \widehat{R}(0) + 2\sum_{k=1}^{\Lambda_n} w(k/\Lambda_n)\widehat{R}(k) \\
&= \tau^2 + o_P(1) + O_P\left(\frac{\log n \log\log n}{n} \right) \left(1 + 2\sum_{k=1}^{\Lambda_n} w(k/\Lambda_n) \right) \\
&= \tau^2 + o_P(1) + O_P\left(\frac{\Lambda_n \log n \log\log n}{n} \right) = \tau^2 + o_P(1).
\end{aligned}$$

The arguments for $\widehat{m} \geqslant m$ are analogous. ■

Now we can prove our main theorem.
Proof of Theorem 3.1. In view of Theorem 3.1 in Hušková and Kirch [6] it suffices to prove

$$P^*\left(\left| \frac{\widehat{d}_n^*}{\widehat{d}_n} - 1 \right| \geqslant \epsilon \right) \to 0 \qquad P-a.s. \tag{5.3}$$

and

$$P^*\left(\left| \frac{\widehat{\tau}^*}{\widehat{\tau}} - 1 \right| \geqslant \epsilon \right) \to 0 \qquad P-a.s. \tag{5.4}$$

We start with proving (5.3). It holds for $\widehat{m} < \widehat{m}^*$

$$\begin{aligned}
&(\widehat{d}_n^* - \widehat{d}_n)/\widehat{d}_n = -\frac{1}{\widehat{d}_n \widehat{m}^*} \sum_{j=1}^{\widehat{m}^*} e^*(j) + \frac{1}{\widehat{d}_n(n-\widehat{m}^*)} \sum_{j=\widehat{m}^*+1}^{n} e^*(j) - \frac{\widehat{m}-\widehat{m}^*}{\widehat{m}^*} \\
&= B_1 + B_2 + B_3.
\end{aligned}$$

First note that by Theorem 2.1 and 3.1 in Hušková and Kirch [6] it holds

$$\frac{\widehat{m}^*}{m} - 1 = \frac{\widehat{m}^* - \widehat{m} + \widehat{m} - m}{m} = O_{P^*}\left(\frac{1}{nd_n^2} \right) + O\left(\frac{\log n}{nd_n^2} \right) = o_{P^*}(1) \qquad P-a.s. \tag{5.5}$$

and analogously

$$B_3 = o_{P^*}(1) \qquad P-a.s.$$

Moreover equation (5.5) together with Lemma 5.4 in [6] implies

$$|B_1| \leqslant \sqrt{\frac{m}{\widehat{m}^*}\frac{1}{m\widehat{d}_n^2}} \max_{1\leqslant j\leqslant n} \frac{1}{\sqrt{j}} \left|\sum_{i=1}^{j} e^*(i)\right| = O_{P^*}\left(\sqrt{\frac{\log n}{d_n^2 n}}\right) = o_{P^*}(1) \qquad P-a.s.,$$

analogously

$$B_2 = o_{P^*}(1) \qquad P-a.s.$$

Analogous arguments for $\widehat{m} \geqslant \widehat{m}^*$ complete the proof of (5.3).

Concerning (5.4) by Theorem 2.1 it suffices to prove

$$P^*(|\widehat{\tau}^* - \tau| \geqslant \epsilon) \to 0 \qquad P-a.s.$$

By Lemma 5.2 in Hušková and Kirch [6] $\operatorname{var}^*\left(\frac{1}{\sqrt{K}}\sum_{k=1}^K e^*(k)\right) = \tau^2 + o(1)$ $P-a.s.$ and by Lemma 5.4 in [6] $\sqrt{K}\bar{e}_n^* = o_{P^*}(1)$ $P-a.s.$, thus it suffices to prove

$$\frac{1}{L}\sum_{l=0}^{L-1}\left(\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(Kl+k)\right)^2 - \operatorname{var}^*\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(Kl+k)\right)\right) = o_{P^*}(1) \qquad P-a.s.$$

From (3.6.8) in Kirch [7] (where the stronger assumption (3.6.6) is not needed because (3.6.9) can be strengthened as pointed out in Hušková and Kirch [6], proof of Lemma 5.2), we obtain ($2 < 2+\widetilde{\delta} \leqslant (2+\delta^{(1)})/(2+\delta^{(2)}+\Delta^{(2)})$, $\widetilde{\delta} < 2$)

$$\begin{aligned}
\mathrm{E}^*\left|\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(k)\right|^{2+\widetilde{\delta}} &= \frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{K}}\sum_{k=1}^{K}\widetilde{e}(j+k)\right|^{2+\widetilde{\delta}} \\
&= O(1)\frac{1}{n}\sum_{j=1}^{n}\left|\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e(j+k)\right|^{2+\widetilde{\delta}} + O(1)\left|\sqrt{K}\bar{e}_n\right|^{2+\widetilde{\delta}} + O(1)\left|\sqrt{K}(d_n+\widehat{d}_n)\right|^{2+\widetilde{\delta}} \\
&= O(1) \qquad P-a.s.
\end{aligned}$$

Since $\left\{\frac{1}{\sqrt{K}}\sum_{k=1}^K e^*(Kl+k) : 1 \leqslant l \leqslant L\right\}$ is conditional (row-wise) i.i.d., an application of the Markov inequality as well as the von Bahr-Esseen inequality (cf. e.g. Shorack and Wellner [11], p. 858) yield

$$\begin{aligned}
&P^*\left(\frac{1}{L}\left|\sum_{l=0}^{L-1}\left(\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(Kl+k)\right)^2 - \operatorname{var}^*\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(Kl+k)\right)\right)\right| \geqslant \epsilon\right) \\
&= O(1)L^{-1-\widetilde{\delta}/2}\,\mathrm{E}^*\left|\sum_{l=0}^{L-1}\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(Kl+k)\right)^2 - \mathrm{E}^*\left(\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(Kl+k)\right)^2\right|^{1+\widetilde{\delta}/2} \\
&= O(1)L^{-\widetilde{\delta}/2}\,\mathrm{E}^*\left|\frac{1}{\sqrt{K}}\sum_{k=1}^{K} e^*(k)\right|^{2+\widetilde{\delta}} = o(1) \qquad P-a.s.
\end{aligned}$$

Thus (5.4) is proven, which completes the proof. ■

Acknowledgement

The authors would like to thank D. Politis (University of California, San Diego) for bringing the relevant papers to our attention.

References

[1] Antoch, J.,, Hušková, M., and Veraverbeke, N. Change-point problem and bootstrap. *J. Nonparametr. Stat.*, 5:123–144, 1995.

[2] Berkes, I., Horváth, L., Kokoszka, P., and Shao, Q.-M. On discriminating between long-range dependence and changes in mean. *Ann. Stat.*, 34(3):1140–1165, 2006.

[3] Berkes, I., Horváth, L., Kokoszka, P., and Shao, Qi-Man. Amost sure convergence of the Bartlett estimator. *Period. Math. Hungar.*, 51:11–25, 2005.

[4] Götze, F. and Künsch, H. R. Second-order correctness of the blockwise bootstrap for stationary observations. *Ann. Stat.*, 24:1914–1933, 1996.

[5] Hušková, M. Permutation principle and bootstrap in change point analysis. *Fields Inst. Commun.*, 44:273–291, 2004.

[6] Hušková, M. and Kirch, C. Bootstrapping confidence intervals for the change-point of time series. *J. Time Ser. Anal.*, 2008. To appear.

[7] Kirch, C. *Resampling Methods for the Change Analysis of Dependent Data.* PhD thesis, University of Cologne, Cologne, 2006. http://kups.ub.uni-koeln.de/volltexte/2006/1795/.

[8] Kirch, C. Block permutation principles for the change analysis of dependent data. *J. Statist. Plann. Inference*, 137:2453–2474, 2007.

[9] Politis, D. N. Adaptive bandwidth choice. *J. Nonparametr. Stat.*, 15:517–533, 2003.

[10] Politis, D. N. and Romano, J. P. Bias-corrected nonparametric spectral estimation. *J. Time Ser. Anal.*, 16:67–103, 1995.

[11] Shorack, G.R., and Wellner, J.A. *Empirical Processes with Applications to Statistics.* Wiley, New York, 1986.

[12] Silverman, B. W. *Density Estimation.* Chapman and Hall, London, 1986.